WANLP 2017
(co-located with EACL 2017)

The Third Arabic Natural Language Processing Workshop

Proceedings of the Workshop

April 3, 2017
Valencia, Spain

©2017 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-945626-44-9

Preface

Welcome to the Third Arabic Natural Language Processing Workshop held at EACL 2017 in Valencia, Spain.

A number of Arabic NLP (or Arabic NLP-related) workshops and conferences have taken place in the last few years, both in the Arab World and in association with international conferences. The Arabic NLP workshop at EACL 2017 follows in the footsteps of these previous efforts to provide a forum for researchers to share and discuss their ongoing work. This particular workshop is the third in a series, following the First Arabic NLP workshop held at EMNLP 2014 in Doha, Qatar; and the Second Arabic NLP workshop held at ACL 2015 in Beijing, China.

We received 47 submissions and selected 22 (47% acceptance rate) for presentation in the workshop. All papers were reviewed by three reviewers on average. The number of submissions is over twice that of the previous workshop in Beijing, which also had a higher acceptance rate (65%). Ten papers will be presented orally and 12 as part of a poster session. The presentation mode is independent of the ranking of the papers. The papers cover a diverse set of topics from Maltese and Arabic dialect processing to models of semantic similarity and credibility analysis, advances in Arabic treebanking, and error annotation for dyslexic texts.

The quantity and quality of the contributions to the workshop are strong indicators that there is a continued need for this kind of dedicated Arabic NLP workshop.

We would like to acknowledge all the hard work of the submitting authors and thank the reviewers for the valuable feedback they provided. We hope these proceedings will serve as a valuable reference for researchers and practitioners in the field of Arabic NLP and NLP in general.

Nizar Habash, General Chair, on behalf of the organizers of the workshop.

Organizers

General Chair
Nizar Habash, New York University Abu Dhabi

Program Chairs
Mona Diab, The George Washington University
Kareem Darwish, Qatar Computing Research Institute
Wassim El-Hajj, American University of Beirut, Lebanon
Hend Al-Khalifa, King Saud University
Houda Bouamor, Carnegie Mellon University in Qatar

Publication Chairs
Nadi Tomeh, LIPN, Université Paris 13, Sorbonne Paris Cité
Mahmoud El-Haj, Lancaster University

Publicity Chairs
Fethi Bougares, University of Le Mans
Wajdi Zaghouani, Carnegie Mellon University-Qatar

Program Committee:

Ahmed Abdelali, Qatar Computing Research Institute, Qatar
Nora Al-Twairesh, King Saud University, Saudi Arabia
Areeb Alowsiheq, Imam University, KSA
Salha Alzahrani, Taif University, Saudi Arabia
Almoataz B. Al-Said, Cairo University, Egypt
Alberto Barrón-Cedeño, Qatar Computing Research Institute, Qatar
Fethi Bougares, Le Mans University, France
Tim Buckwalter, University of Maryland, USA
Violetta Cavalli-Sforza, Al Akhawayn University, Morocco
Abeer Dayel, King Saud University, Saudi Arabia
Tamer Elsayed, Qatar University, Qatar
Ossama Emam, IBM, USA
Ramy Eskander, Columbia University, USA
Nizar Habash, New York University Abu Dhabi, UAE
Bassam Haddad, University of Petra, Jordan
Hazem Hajj, American University of Beirut, Lebanon
Maha Jarallah Althobaiti, Taif University, Saudi Arabia
Azzeddine Mazroui, University Mohamed I, Morocco
Karine Megerdoomian, The MITRE Corporation, USA
Ghassan Mourad, Université Libanaise, Lebanon
Hamdy Mubarak, Qatar Computing Research Institute, Qatar
Preslav Nakov, Qatar Computing Research Institute, Qatar

Alexis Nasr, University of Marseille, France
Kemal Oflazer, Carnegie Mellon University Qatar, Qatar
Eshrag Refaee, Jazan University, Saudi Arabia
Mohammad Salameh, Carnegie Mellon University, Qatar
Hassan Sawaf, eBay Inc., USA
Khaled Shaalan, The British University in Dubai, UAE
Khaled Shaban, Qatar University, Qatar
Otakar Smrž, Džám-e Džam Language Institute, Czech Republic
Wajdi Zaghouani, Carnegie Mellon University, Qatar
Imed Zitouni, Microsoft Research, USA

Invited Speaker:

Stephan Vogel, Qatar Computing Research Institute (QCRI), Qatar

Table of Contents

Workshop Program

Monday April 3, 2017

09:00–11:30 Session AAA: Session

09:00–09:10 *Opening Remarks*
Nizar Habash

09:10–10:00 *Keynote*
Stephan Vogel

10:00–10:20 *Identification of Languages in Algerian Arabic Multilingual Documents*
Wafia Adouane and Simon Dobnik

10:20–10:40 *Arabic Diacritization: Stats, Rules, and Hacks*
Kareem Darwish, Hamdy Mubarak and Ahmed Abdelali

10:40–11:00 *Semantic Similarity of Arabic Sentences with Word Embeddings*
El Moatez Billah Nagoudi and Didier Schwab

11:00–11:30 ***Coffee Break***

11:30–12:50 Session BBB: Session

11:30–11:50 *Morphological Analysis for the Maltese Language: The challenges of a hybrid system*
Claudia Borg and Albert Gatt

11:50–12:10 *A Morphological Analyzer for Gulf Arabic Verbs*
Salam Khalifa, Sara Hassan and Nizar Habash

12:10–12:30 *A Neural Architecture for Dialectal Arabic Segmentation*
Younes Samih, Mohammed Attia, Mohamed Eldesouki, Ahmed Abdelali, Hamdy Mubarak, Laura Kallmeyer and Kareem Darwish

12:30–12:50 *Sentiment Analysis of Tunisian Dialects: Linguistic Ressources and Experiments*
Salima Medhaffar, Fethi Bougares, Yannick Estève and Lamia Hadrich-Belguith

12:50–14:30 ***Lunch Break***

Monday April 3, 2017 (continued)

14:30–16:30 Session CCC: Session

14:30–14:50 *CAT: Credibility Analysis of Arabic Content on Twitter*
Rim El Ballouli, Wassim El-Hajj, Ahmad Ghandour, Shady Elbassuoni, Hazem Hajj and Khaled Shaban

14:50–15:10 *A New Error Annotation for Dyslexic texts in Arabic*
Maha Alamri and William J Teahan

15:10–15:30 *An Unsupervised Speaker Clustering Technique based on SOM and I-vectors for Speech Recognition Systems*
Hany Ahmed, Mohamed Elaraby, Abdullah M. Mousa, Mostafa Elhosiny, Sherif Abdou and Mohsen Rashwan

15:30–16:00 ***Poster Boaster (2 min each for 12 papers)***

16:00–16:30 ***Coffee Break***

16:30–18:00 Session DDD: Session

SHAKKIL: An Automatic Diacritization System for Modern Standard Arabic Texts
Amany Fashwan and Sameh Alansary

Arabic Tweets Treebanking and Parsing: A Bootstrapping Approach
Fahad Albogamy, Allan Ramsay and Hanady Ahmed

Identifying Effective Translations for Cross-lingual Arabic-to-English User-generated Speech Search
Ahmad Khwileh, Haithem Afli, Gareth Jones and Andy Way

A Characterization Study of Arabic Twitter Data with a Benchmarking for State-of-the-Art Opinion Mining Models
Ramy Baly, Gilbert Badaro, Georges El-Khoury, Rawan Moukalled, Rita Aoun, Hazem Hajj, Wassim El-Hajj, Nizar Habash and Khaled Shaban

Robust Dictionary Lookup in Multiple Noisy Orthographies
Lingliang Zhang, Nizar Habash and Godfried Toussaint

Arabic POS Tagging: Don't Abandon Feature Engineering Just Yet
Kareem Darwish, Hamdy Mubarak, Ahmed Abdelali and Mohamed Eldesouki

Monday April 3, 2017 (continued)

Toward a Web-based Speech Corpus for Algerian Dialectal Arabic Varieties
Soumia Bougrine, Aicha Chorana, Abdallah Lakhdari and Hadda Cherroun

Not All Segments are Created Equal: Syntactically Motivated Sentiment Analysis in Lexical Space
Muhammad Abdul-Mageed

An enhanced automatic speech recognition system for Arabic
Mohamed Amine Menacer, Odile Mella, Dominique Fohr, Denis Jouvet, David Langlois and Kamel Smaili

Universal Dependencies for Arabic
Dima Taji, Nizar Habash and Daniel Zeman

A Layered Language Model based Hybrid Approach to Automatic Full Diacritization of Arabic
Mohamed Al-Badrashiny, Abdelati Hawwari and Mona Diab

Arabic Textual Entailment with Word Embeddings
Nada Almarwani and Mona Diab